

	QMRF identifier (JRC Inventory): Q13-410-0036
	QMRF Title: QSAR for haloacetic acid mutagenicity
	Printing Date: Dec 11, 2019

1. QSAR identifier

1.1. QSAR identifier (title):

QSAR for haloacetic acid mutagenicity

1.2. Other related models:

1.3. Software coding the model:

2. General information

2.1. Date of QMRF:

29 September 2009

2.2. QMRF author(s) and contact details:

[1] Alfonso Pérez-Garrido Environmental Engineering and Toxicology Department, Catholic University of San Antonio, Guadalupe, Murcia, C.P. 30107, Spain Aperez@pdi.ucam.edu

[2] Maykel Pérez González Environmental Engineering and Toxicology Department, Catholic University of San Antonio, Guadalupe, Murcia, C.P. 30107, Spain

[3] Amalio Garrido Escudero Environmental Engineering and Toxicology Department, Catholic University of San Antonio, Guadalupe, Murcia, C.P. 30107, Spain

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Alfonso Pérez-Garrido Environmental Engineering and Toxicology Department, Catholic University of San Antonio, Guadalupe, Murcia, C.P. 30107, Spain Aperez@pdi.ucam.edu

2.6. Date of model development and/or publication:

The model was published in 2008 (Bioorganic & Medicinal Chemistry 16, 5720-5732).

2.7. Reference(s) to main scientific papers and/or software package:

Pérez-Garrido A, González MP & Escudero AG (2008). Halogenated derivatives QSAR model using spectral moments to predict haloacetic acids (HAA) mutagenicity. Bioorganic & Medicinal Chemistry 16, 5720–5732.

2.8. Availability of information about the model:

Non-proprietary

2.9. Availability of another QMRF for exactly the same model:

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Salmonella typhimurium (TA100 strain)

3.2. Endpoint:

4. Human Health Effects 4.10. Mutagenicity

3.3. Comment on endpoint:

The activity is defined as the logarithm of TA100 strain *S. typhimurium* Ames test without activation (S9) and with preincubation. It is calculated as the slope of the linear portion of the dose–response curve.

3.4.Endpoint units:

reversions/nmol

3.5.Dependent variable:

log (reversions/nmol), log TA100

3.6.Experimental protocol:

Ames test (Mortelmans and Zeiger, 2000) TA100 strain without activation and with preincubation.

3.7.Endpoint data quality and variability:

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Multilinear regression QSAR

$$\log \text{TA100} = 8.95 \times 10^{-11} \text{Std} - 5.77 \times 10^{-2} \text{Pols} + 5.87 \times 10^{-8} \text{Mol} - 8.57 \times 10^{-10} \text{Dip} + 2.3 \times 10^{-8} \text{Pols} + 1.75 \times 10^{-2} \text{Dip2} - 4.07 \times 10^{-10} \text{Dip2} - 2.26 \times 10^{-8} \text{Pol} - 1.24$$

4.3.Descriptors in the model:

TOPS-MODE It based on the calculation of the spectral moments of the so-called bond matrix (Estrada, 1996 and 1997).

4.4.Descriptor selection:

The variables to be included in the equation were selected using the forward stepwise procedure as variable selection strategy (Draper and Smith, 1981).

4.5.Algorithm and descriptor generation:

The spectral moments of the edge adjacency matrix are defined as the traces. That is the sum of the main diagonal of the different powers of such matrix. Several bond weights such as standard bond distance (Std), standard bond dipole moments (Dip, Dip2), hydrophobicity (H), polar surface area (Pols), polarizability (Pol), molar refractivity (Mol), van der Waals radii (vdW), and Gasteiger–Marsilli charges (Gas) were used for computing the spectral moments of the bond matrix.

4.6.Software name and version for descriptor generation:

Modeslab

<http://www.modeslab.com/>

4.7.Chemicals/Descriptors ratio:

(40 chemicals/ 8 descriptors) = 5

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain is established inside a squared area (of the Williams plot) within ± 2 standard deviations and a leverage threshold h^* of 0.6 ($h^* = 3j/n$, being j the number of model parameters and n the number of objects).

5.2.Method used to assess the applicability domain:

Method based on leverage values (Gramatica, 2007)

5.3.Software name and version for applicability domain assessment:

Mobydigs Computer Software 1.0

Talete srl

http://www.talete.mi.it/products/moby_description.htm

5.4.Limits of applicability:

Compounds with a less leverage value than leverage threshold h^* of 0.6

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS RN: Yes

Chemical Name: Yes

Smiles: No

Formula: No

INChI: No

MOL file: No

6.3.Data for each descriptor variable for the training set:

All

6.4.Data for the dependent variable for the training set:

All

6.5.Other information about the training set:

40 chemicals

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

$R^2= 0.902$ (Coefficient of determination);

$s=0.548$ (Standard error);

$F=35.730$ (Fisher function);

$AIC= 0.727$ (Aikake Information Criterion);

$FIT=1.529$ (Kubinyi function)

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

The model was evaluated by leave-one-out cross validation.

$Q^2(LOO)=0.842$

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

6.10.Robustness - Statistics obtained by Y-scrambling:

$a(r^2)=0.125$;

$a(Q^2)=-0.392$

6.11.Robustness - Statistics obtained by bootstrap:

$Q^2_{boot}=0.718$

6.12.Robustness - Statistics obtained by other methods:

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

No

7.2. Available information for the external validation set:

CAS RN: No

Chemical Name: No

Smiles: No

Formula: No

INChI: No

MOL file: No

7.3. Data for each descriptor variable for the external validation set:

No

7.4. Data for the dependent variable for the external validation set:

No

7.5. Other information about the external validation set:

7.6. Experimental design of test set:

7.7. Predictivity - Statistics obtained by external validation:

7.8. Predictivity - Assessment of the external validation set:

7.9. Comments on the external validation of the model:

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

Mechanistic interpretation is given in a published paper (Pérez-Garrido et al., 2008)

8.2. A priori or a posteriori mechanistic interpretation:

A posteriori interpretation.

8.3. Other information about the mechanistic interpretation:

9. Miscellaneous information

9.1. Comments:

9.2. Bibliography:

[1]Mortelmans K & Zeiger E (2000). The Ames Salmonella/microsome mutagenicity assay. *Mutation Research* 455, 29– 60.

[2]Draper NR & Smith H (1981). *Applied Regression Analysis*, 2nd ed. John Wiley and Sons, New York.

[3]Gramatica P (2007). Principles of QSAR models validation: internal and external. *QSAR & Combinatorial Science* 26, 694-701.

[4]Pérez-Garrido A, González MP & Escudero AG (2008). Halogenated derivatives QSAR model using spectral moments to predict haloacetic acids (HAA) mutagenicity. *Bioorganic & Medicinal Chemistry* 16, 5720–5732.

[5]Estrada E (1996). Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. 1. Definition and Applications to the Prediction of Physical Properties of Alkanes. *Journal of Chemical Information and Computer Sciences* 36, 844–849.

[6]Estrada E (1997). Spectral Moments of the Edge-Adjacency Matrix of Molecular Graphs. 2. Molecules Containing Heteroatoms and QSAR Applications. *Journal of Chemical Information and Computer Sciences* 37, 320–328.

9.3. Supporting information:

HAA training_40.sdf

<http://qsar.db.jrc.ec.europa.eu/qmrf/protocol/Q13-410-0036/attachment/A698>

Test set(s) Supporting information

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

Q13-410-0036

10.2. Publication date:

2013-06-27

10.3. Keywords:

Ames mutagenicity; TOPS-MODE; haloacetic acid;

10.4. Comments:

former Q8-26-8-155